SYNOPSIS

"Effective Multimedia Framework to Detect and Diagnose Breast Cancer Using Machine Learning"

A Synopsis submitted to

SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY, Nanded

For the Award of Degree

DOCTOR OF PHILOSOPHY

In

Computer Sciences

By

Mr. Riyadh Mohammed Musaed Al-Tam

UNDER THE GUIDANCE OF

Dr. Sachin M. Narangale

School of Computational Sciences

2020

Introduction

Breast Cancer is a disease that occurs in the glandular epithelium of the breast and afflicts a significant number of humans worldwide. Whenever a suspicion is raised, periodical exams usually including digital mammograms (DM), ultrasound (US), magnetic resonance imaging (MRI), microscopic (histological) images, Infrared thermography, or other tools might be recommended [3]. Therefore, early detection of breast cancer will increase the survival rate and help for early choosing the right treatment. Mammography is the most popular tool used to detect cancer, however, the efficient way to improve early detection accuracy is to combine different imaging methods such as x-ray (mammography), ultrasound and magnetic resonance imaging jointly. With modern technologies in the field of medicine, a huge number of medical imaging have been collected and are available to the medical research community. However, physicians have one of the most interesting and challenging tasks in getting an accurate prediction of a disease outcome and suggest a suitable treatment [3]. Detection of breast cancer using mammography modality is challenging as an image classification task because of that the tumors have only a small portion of the image of the entire breast. For example, a cancerous region of interest (ROI) from a full-field digital mammography (FFDM) image can be as small as 100×100 pixels and the entire image is typically 4000×3000 pixels. Therefore, many studies have limited their focus to classify only annotated images since subtle differences in disease states are more difficult to perceive than differences in everyday objects [2]. To improve the accuracy of deep learning algorithms to success in predicting breast cancer, well-annotated large data sets are needed since deep networks are complex, computer software and hardware are evolving constantly. Region-based convolutional neural network (R-CNN) and its variants can achieve good results if mammogram images are already manually annotated ROIs and can be readily applied [2]. However, most of the mammography databases that already exist contain images without any annotations. For this reason, we must find a way to balance using manually annotated datasets with other datasets that are not annotated to achieve good results. The methods that require annotated ROIs often cannot be applied to large mammography databases that lack ROI annotations. The problem of annotated ROI of breast cancer images is that they are applying in a laboratory by radiologists and costly to assemble. Another trend of studies has attempted to train neural networks using whole mammograms without relying on annotated

images. It might take a longer time to locate lesions on the corresponding portions of the mammogram images.

Machine learning is the scientific area that makes computers to learn from data [5]. The machine learning techniques have proven to play an important role in detecting cancer in medical images and can be used to suggest a suitable treatment according to the type of cancer [4]. These techniques can discover and identify patterns and relationships between them, from complex data sets, while they are able to effectively predict future outcomes of a cancer type. Various machine learning algorithms have been used across decades, for instance, linear discriminant analysis, support vector machines, decision trees and random forests, and neural networks techniques. There are three main common types of ML methods known as (I) supervised learning, (II) unsupervised learning, and (III) reinforcement learning [1]. In supervised learning, a labeled set of training data is used to estimate or map the input data to the desired output. In contrast, under the unsupervised learning methods, no labeled examples are provided and there is no notion of the output during the learning process. As a result, it is up to the learning scheme/model to find patterns or discover the groups of the input data. In supervised learning, this procedure can be thought of as a classification problem [1]. The task of classification refers to a learning process that categorizes the data into a set of finite classes. Two other common ML tasks are regression and clustering. In the case of regression problems, a learning function maps the data into a real-value variable. Subsequently, for each new sample, the value of a predictive variable can be estimated, based on this process [1]. Clustering is a common unsupervised task in which one tries to find the categories or clusters in order to describe the data items. Based on this process each new sample can be assigned to one of the identified clusters concerning the similar characteristics that they share. The reinforcement learning where an agent learns how to behave in an environment by performing an action and seeing the results. Supervised learning is the most type of machine learning applied to detect cancer.

In this study, a final application to detect the location of breast cancer using some of the best used machine learning algorithms will be created. Besides, the ability of the application to suggest the suitable treatment that should be supported according to the current status of breast image. Different accurate datasets of the breast cancer will be collected to support and enhance the detection of breast cancer in both phases: training and testing phases. Moreover, check if the collected datasets have some required features of suitable treatment for each image, or new

features with the suitable treatment will be added to the collected datasets. The new added features should be revised by helping experts of the domain and this might need a long time to be done. The final application should be accessible by radiologists anywhere over the world.

Literature Survey

Kumar, Aman, and M. Poonkodi [18] conducted a comparison on the basis of confusion matrix and accuracy for breast cancer diagnosis between different supervised learning classification algorithms such as logistic regression, KNN classifier, SVM, naïve Bayes, decision trees. The suggested algorithm, namely, Kernel SVM with PCA achieves 98.24% accuracy as compared to others. 97.36% accuracy is be given by using Fusion of KNN, SVM, and naïve Bayes by stacking to be combined using logistic regression.

Khan, A. A., and A. S. Arora [29] explores if thermograph can be a better alternative to the standard mammography which is being a painful procedure and exposure of the body to harmful X-rays. Thermography is coming out technique with advantages of being non-invasive, safe, portability and cost effectiveness. Some of thermograms images are taken from DMR-Database for Mastology Research. Gabor filters is used to extract the texture features of the left and right breasts, support vector machine (SVM) is used to classify breast lesions based on the textural asymmetry between the breasts into normal and cancerous cases. The proposed methodology attained 84.5% of accuracy using Gabor features and SVM classifier, therefore, thermography is able to detect early breast cancer.

Some recently proposed methods are investigated for breast cancer detection using computer vision and machine learning techniques by Lu, Yao[34]. The detection performance of different methods on mammography images and histological images are compared and analyzed. To enhance the diagnosis of breast cancer, three imaging modalities are used: histological imaging, x-ray (mammography) imaging and ultrasound imaging. Different datasets already exist to each type of modality. This paper shows that deep learning based methods have achieved good results for breast cancer detection using histological imaging or mammography imaging, specially the ScanNet has achieved the highest accuracy on histological images. Besides, impressive result could be achieved using Adaboost + Random Forest classifier on mammography images.

Lu, Hao-Chun, El-Wui Loh, and Shih-Chen Huang [35] used deep learning techniques to detect and diagnose breast cancer using 9000 mammograms. The convolutional neural network is applied to train a classified model, the preprocessing techniques are implemented using median filter, contrast-limited adaptive histogram equalization, and data augmentation. The final outcomes shows that the model with preprocessed images achieves higher accuracy than the model without preprocessed images. 70% of the dataset is chosen as training data, 10% of dataset is selected as validation data, and the rest of the dataset works as testing data. BI-RADS, is breast cancer score to distinguish benign and cancer cases using number start from 1 to 6. The overall sensitivity, specificity, and BI-RADs score achieved by the proposed model with preprocessed images were 0.91, 0.57, 0.88, while they were 0.79, 0, 0.88, using the proposed model without reprocessed images.

Yassin, Nisreen [48] presented a systematic review to investigate the state of the art regarding the computer aided diagnosis/detection (CAD) systems for breast cancer. Popular scientific databases are used like Springer Link (SL), Science Direct (SD), IEEE Xplore Digital Library, and PubMed. Inclusion and exclusion criteria were defined and applied to select only 154 studies from 320 studies retrieved. This paper was limited to scientific and academic works and excluded commercial interests.

Amin Zadeh Shirazi [49] used unsupervised and supervised learning techniques such as self-organizing map (SOM) and complex-valued neural network (CVNN) to create a hybrid computational intelligence model for reliable detection of breast cancer. The dataset consists of 822 input samples each with five features (mass shape, mass density, mass margin, age of patient, and Breast Imaging Reporting and Data System assessment). The proposed model was divided into two steps. In the first step, considering the input features, patient's features with the most similarity were clustered by using SOM technique. In the second step, for each cluster, breast cancer images is classified into benign or malign using complex-valued neural network. Receiver operating characteristic analyses and confusion matrix is used to compare the outcomes of the current proposed model with the medical diagnosis results. As indicated by the test dataset confusion matrix analysis, it is noticed that 95% is the ratio of distinguishing the sick from the true sick, the ability to distinguish the healthy from the true healthy is 94.2%, and the proportion of the total number of correct diagnosis of the test as sick and healthy is 94.5%. In like manner,

the superiority of the proposed model was confirmed and can be used for reliable and robust detection of breast cancer.

Vasquez, JA Tobon [47] presented a study for using microwave imaging (MWI) system prototype to early detect breast cancer. In the built MWI prototype the vector network analyzer (VNA) is replaced with low cost and small form-factor COTS. In this work, an experiment is done using frequency synthesizer (a transmitter, TX) and a Direct Conversion (DC) receiver (RX) as the two main components of the system. DC RX is selected since it's a strong in phase mismatches and phase noise. Two custom wideband monopole antennas are used to connect the transmitter and the receiver together. The antennas, together with the breast phantom, are inundated in a coupling liquid (a mixture of glycerin and water) to mirror the normal breast tissue so as to diminish the mismatch because of the interface between the antennas and the breast. The phantom is comprised by a plastic cylinder with 7 mm diameter and filled with water to act the malignant tissue. The 7-mm cylinder is inserted in a larger plastic cylinder with 8 cm diameter filled with a mixture of glycerin and water (80-20) that acts as breast fat tissue. Interferometric Multiple Signal Classification (I-MUSIC) algorithm, is used to elaborate the measured data, is an example of a linear method robust and computationally effective. In particular, the angle between the RX and TX antennas is equal to 135 degree, automated antenna rotation around the breast is done. The complete time for one measurement set and data elaboration is overwhelmed by the rotation time of the system which is about 1.5 minute. Great detections are accounted for straightforward breast phantoms.

Two-hybrid feature selection methods (SFSP and SBSP) are presented to be utilized in the diagnosis of breast cancer fast and effectively by Uzer, Mustafa Serter, Onur Inan, and Nihat Yılmaz [46]. SFSP and SBSP methods which are made by consolidating the sequential forward selection (SFS) and the sequential backward selection (SBS) together with the principal component analysis (PCA) created by using quadratic discriminant analysis classification (QDA) algorithmic criteria. Besides, the tenfold cross-validation method is used as criteria during the selection of the features in this work. The proposed two-hybrid feature methods are helped in reducing of the dimension of the feature space for input from 9 to 4. Artificial Neural Networks is utilized as a classifier. To increase the reliability of the outcomes, the cross-validation method has been favored additionally in the phase of the classification of the proposed system. To some extent determine the correctness of the proposed system, the Wisconsin Breast Cancer Database

obtained from the UCI is utilized which contains from 699 records of patients. The average correctness of the classification in the gained outcomes are 97.57% for SFSP + NN and 98.57 % for SBSP + NN.

Wang, Xin [45] presented a study aiming to build a mathematical nomogram to predict breast cancer survival and guide postoperative adjuvant chemotherapy in the Republic of China. A sample with 5,504 breast cancer patients from the Tianjin Breast Cancer Cases Cohort was collected. To investigate the factors associated with overall survival (OS), multivariable Cox regression was used. Under the curve (AUC) and the calibration curve are used to evaluate the performance of the suggested nomogram. A partial score of death was classified into three subgroups: low, moderate and high-risk subgroups. The proposed nomogram includes different factors like Age, diameter, histological differentiation, lymph node metastasis, estrogen, and progesterone receptor. The nomogram was not only able to predict the 3-year overall survival [AUC =74.1%; 95% confidence interval (CI): 70.1–78.0%] but also was able to predict 5-year overall survival [AUC =72.3%; 95% CI: 69.6–75.1%]. In high risk subgroup, adjuvant chemotherapy was negatively associated with death [Hazard Ratio (HR) = 0.54; 95% CI: 0.37–0.77; P<0.001]. In spite of that, there is no significant association was discovered in groups with low (HR=1.47; 95% CI: 0.52–4.19; P=0.47) and moderate risk (HR=0.78; 95% CI: 0.42–1.48; P=0.45). To perform balancing in observed covariates between patients received and not received adjuvant chemotherapy, the propensity score matching (PSM) method was used. Kaplan-Meier's method indicated that high-risk patients could benefit from chemotherapy, while the benefit from chemotherapy in low risk and moderate risk patients did not seem to appear. By and large, the proposed nomogram could be utilized to assess the overall survival of breast cancer patients and anticipate the extent of advantage and guide adjuvant chemotherapy for breast cancer patients after surgery.

Takada, Masahiro [44] used a machine learning to predict human epidermal growth factor receptor 2 (HER2)-positive breast cancer of patients who were treated using neoadjuvant chemotherapy (NAC) plus trastuzumab drugs at the duration between 2001 and 2010. This work can be used to help optimize adjuvant therapy and postoperative surveillance. 776 patients data from multicenter were collected. Two models were developed using machine learning namely, disease-free survival (DFS) (DFS model) and brain metastasis (BM) (BM model), to predict of recurrence of disease within 5 years after surgery. Results for DFS model and BM model using

area under the curve (AUC) values were 0.785 (95% CI 0.740–0.831, P < 0.001) and 0.871 (95% CI 0.830–0.912, P < 0.001) respectively. The present study shows that low-risk DFS or BM events of patients gives better 5-year DFS and BM rates (89% for DFS and 99% for BM models) compares to those with high-risk DFS or BM events (61% for the DFS model and 87% for the BM model).

Syed, Liyakathunisa, Saima Jabeen, and S. Manimala [43] presented a study for showing the importance of information and communication technologies (ICT) and Internet of Things (IoT) to deliver healthcare services to remote, distant, and underserved regions and to early detect breast cancer using mammographic images. In general, the advances in telehealth services are presented and also a novel system for early detection of breast cancer from remote and underserved areas is proposed. To help people in remote and rural locations to detect breast cancer, telemammography is used, which is one of the most publicized telehealth services. Besides, to enhance the resolution of used mammographic images in this system, efficient wavelet-based image processing techniques for preprocessing detection is applied. Different classification algorithms are implemented to select the best classification algorithms. As well as, Multi-Layer Perceptron Neural Networks, J48 decision trees, Random Forest, and K-Nearest Neighbor classifiers are used for classifying the tumor into three categories namely: benign, malignant, and normal. The classification in this work depends on the area, volume, and boundaries of tumor masses. Moreover, all the cancer features and classification methods are compared using accuracy, sensitivity, specificity, precision, and mean square error measurements. Mammographic Image Analysis Society (MIAS) database is used in the proposed system. The final results using neural network classifier outperforms the results given by using K-NN, random forest, and decision trees classifiers.

Deep neural network is used to detect breast cancer using microwave imaging technique by Suresh, Veena, and R. Kiran [42]. Microwave modality is a new technique that can be used for imaging the breast of human body, and it is safer than mammography and MRI, as it is nonionizing, non invasive and economical. In microwave region, cancerous regions can be detected by using the dielectric properties of breast tissue. Scattered fields are produced when breast tissue is imaging using the microwave modality. In this paper, to detect cancerous voxels, finding the complex permittivity from the scattered field is done. Each malignance has different structure with constant dielectric properties that can be used as input to train the deep learning

network and then to distinguish between benign, malignant and skin voxels. The proposed technique achieves 99.16% for the malignancy of shape 4x4x4, 96.43% for the malignancy of shape 4x4x3, 97.03% for the malignancy of shape 4x3x4, and 96.82% for the malignancy of shape 3x4x4. Therefore, a challenging task still there to detect breast cancer as the size of malignancy decreases.

Michahial, Stafford, and Bindu A [36] applied the hybridization of cuckoo search algorithm with the multiple kernel-based fuzzy c means (MKF-CS) method and hybrid-based neural (H-BN) classifier on a group of ultrasound images for detection of breast cancer. MKF-CS algorithm is used to extract ROI from input images before the stage of diagnosis of breast cancer. Some of the main features are extracted and fed to the H-BN classifier such as mean, variance, standard deviation, and so on. For the detection of breast cancer, an H-BN classifier is applied to classify the data into benign or malignant tumor classes. Accuracy, sensitivity, and specificity metrics are utilized to evaluate the performance of the suggested approach with values of 0.8889, 1, and 0.85 respectively.

Song, Hongchao, Yunpeng Li, and Aidong Men [41] used machine learning techniques to detect breast cancer using microwave images. Feature extraction methods based on time-frequency representations are applied, including the wavelet transform and the empirical mode decomposition (EMD). The discrete wavelet transform (DWT), the dua-tree complex wavelet transform (DTCWT) and the empirical mode decomposition (EMD) methods are investigated. To reduce the misalignment of data and to provide features of data with more robust, time invariant statistics can be generated. Final results show that the detection performance of breast cancer is improved by using the proposed methods with an ensemble selection-based classifier.

Shen, Dinggang, Guorong Wu, and Heung-Il Suk [40] introduced and reviewed commonly used deep learning methods for breast cancer detection and diagnosis. Even if the deep learning methods achieve high performance across different medical applications; however, there is still a place for improvement. First, a dataset with millions of annotated images is required to enhance the output of deep learning methods. Second, to enhance the accuracy of the output of deep learning methods, data-driven feature representations are used, especially in unsupervised learning, but it would be desirable to devise a new methodological architecture involving domain-specific knowledge. Third, it is necessary to develop techniques to efficiently detect and diagnose cancer using images acquired from different scanning modalities.

Sharma, Shubham, Archit Aggarwal, and Tanupriya Choudhury [39] reviewed the largely popular machine learning techniques commonly used for breast cancer prediction, namely Random Forest, kNN (k-Nearest-Neighbor) and Naïve Bayes. The Wisconsin Diagnosis Breast Cancer data set is used to test the proposed methods. Two popular measurements are used : accuracy and precision measurements. According to the final results, kNN had the best accuracy, precision and F1 score over the other algorithms with values 95.90, 98.27,94.20 respectively.

Selvathi, D., and A. Aarthy Poornila [38] used deep learning techniques such as convolutional neural network, sparse autoencoder, and stacked sparse autoencoder to early detect breast cancer using mammogram with the minimal errors. To test the suggested methods, the mini-MIAS mammographic database is used which contains 322 digitized films of 1024 x 1024 images. The stacked sparse autoencoder performs better compared to other suggested methods with 98.9% of accuracy, 97.83% of sensitivity, 99.38% of specificity, 98.98 of prevision, and 99.56 of negative predictive value.

A novel ensembling method is proposed, namely CFDNN, to reduce the computational expense in the diagnosis of breast cancer by Sanwal, Karan, and Himanshu Ahuja [37]. To enhance the performance of suggested CFDNN, and by taking advantages of a multiprocessor-based architecture, a parallelizable global structure is applied. The Wisconsin Breast Cancer dataset is used in this paper. According to the writers, with regard to the computational aspect, the proposed CFDNN model achieves the fastest model in breast cancer detection among other state-of-the-art methods with an accuracy of 99.01%. Therefore, CFDNN can be used as an optimal choice for providing a second opinion in medical diagnosis.

Král, Pavel, and Ladislav Lenc [33], a novel method namely Local Binary Patterns (LBP), is used for breast cancer detection from mammography images. Some of images are extracted from MIAS and DDSM databases and used as input the proposed method. This method is composed of the four main steps : image preprocessing, local binary patterns computation, image representation creation, and cancer detection. The overall accuracy by the suggested method is about 84%.

Kourou, Konstantina [32], recent of machine learning methods are reviewed to detect cancer progression. The supervised ML techniques are used in predictive models. The most recent papers that employ ML techniques are mentioned and used to predict cancer risk or patient outcomes. Most of the reviewed studies use different types of input data like : genomic,

clinical, histological, imaging, demographic, epidemiological data or combination of these. Some of the algorithms used with oral, lung cancer or other type of cancer but the ones that used with breast cancer are Artificial Neural Network (ANN), Graph-based Semi-supervised Learning (SSL),SVM, Bayesian Networks (BN), Decision Tree (DT), SSL Co-training algorithm. Delen D et al. apply DT method on 200000 patient images, the type of data is SEER cancer database ,and achieves 93% accuracy. Eshlaghy A et al. apply SVM method on 547 images, the type of data is Clinical or population, and achieves 95% accuracy. Ayer T et al. apply ANN method on 62219 patient images, the type of data input is Mammographic or demographic, 0.965 of accuracy is attained.

Kontos, Konstantinos, and Manolis Maragoudakis [31] proposed using data mining techniques to detect breast cancer with mammography images. Traditional data mining algorithms faces a several challenges due to the high dimensionality and class imbalance of training data. Therefore, genetic algorithms are applied to reduce the feature set to the informative ones while a hybrid boosting and genetic sub-sampling approach are used to solve the problems of class imbalance issues. Decision Trees method is applied for feature extraction. Adaboost and k-Nearest Neighbor achieves the best precision and recall rates with an overall accuracy 99.24% is reached using a dataset of 45 mammograms.

Khourdifi, Youness, and Mohamed Bahaj [30], applying the best four machine learning algorithms are conducted to early predict breast cancer including Random Forest, Naïve Bayes, Support Vector Machines SVM, and K-Nearest Neighbors K-NN. In this paper, the experimental results show that SVM is superior the other methods and achieves the highest accuracy 97.9%.

Jalalian, Afsaneh [26] reviewed some of the most recent algorithms to help in detect and diagnosis breast cancer using mammography and ultrasound images. To increase the accuracy of diagnosis, computer-aided detection or diagnosis (CAD) systems have been used. A CAD system in general composed of four stages: preprocessing, segmentation of regions of interest, feature extraction and selection, and finally classification. Some of image's features that might help to detect or diagnosis lesions are, sensitivity, size of lesions, tissue density, histopathology of lesions, and the number of false-positive values per image. In this reviewing, the number of images inside the dataset used by each paper is different, however, the writer shows that the proposed algorithm of Sahiner achieved the highest performance among the others for mass diagnosis methods with value equal 94%. In addition to, Kallergi attains 0.98% for calcification

diagnosis methods. Sahiner and Kallergi apply their algorithms using mammography images. For Ultrasound images, Huang and Chen achieves the highest performance with value equal 97%.

R.Chtihrakkannan, P.Kavitha, T.Mangayarkarasi, R.Karthikeyan [27] enhanced an image processing algorithm to early detect breast cancer. The mammography images used as input and Gaussian Filter and Edge detection techniques are applied to enhance image quality. Wavelet transform and Grey-Level Co-occurrence Matrix (GLCM) are implemented for feature extraction. Deep neural networks (DNN) is a Multilayer supervised classifier, is applied in this work. In training stage, dataset images are created, while, in testing stage, the acquired image from a patient is processed first using image processing techniques such as pre-processing and feature extraction, and then is processed by the DNN classifier to get the final outputs of the system : normal and abnormal cases. Python language is used for implementing the proposed system. Some of algorithms are reviewed like artificial neural network, support vector machine, k-nearest neighbors, and DNN which superior the other and gives the highest accuracy with value of 92%.

The rate of false-positive in breast cancer diagnosis has been decreased using the combination of a conventional US and Color Doppler by Keatmanee, Chadaporn [28]. Ultrasound (US) images is used in this work and a novel algorithm is proposed to automatically initialize an active model by utilizing the fusion of conventional US and Color Doppler. Examples and comparisons with three state-of-the-art automatic initialization methods including center of divergence (COD), field segmentation (FFS), Poisson inverse gradient (PIG), are explained to show the superiority of the proposed algorithm. The proposed method gains higher accuracy and fast convergence to features of interest. The improving the accuracy of the suggested method is under investigation using datasets contain few vascular flows in the Color Doppler.

Guzmán-Cabrera [20] suggested an algorithm to early detect breast cancer using several mammographies images that are taken from the digital database. It depend on three steps to differentiate masses and microcalcifications from the background tissue, morphological operators are used, machine learning techniques are applied to extract masses and microcalcifications cases and a clustering algorithm for intensity-based segmentation. The DDSM database are used which contains a group of images that already analyzed, classified, and hand-labeled.

Aydin, Emine Avşar, and Mümine Kaya Keleş [22] used a rectangular microstrip patch antenna (RMPA) for microwave breast cancer detection. One of the data mining methods is used, namely the random forest algorithm. According to the results of this work, the random forest method can detect breast cancer tumors with a value of 94% of accuracy.

Aydin, Emine Avşar, and Mümine Kaya Keleş [23] proposed Curvelet And probabilistic neural network (PNN) Classifier to early detect breast cancer. For feature extraction, a curvelet and texture analysis are applied, while for classification, a PNN classifier is used. Mini Mammographic Image Analysis Society(Mini-MIAS) dataset is used and the overall accuracy of this work has 93%.

An effect method was conducted by Ghasemzadeh, Ardalan, Saeed Sarbazi Azad, and Elham Esmaeili to classify and detect breast cancer using mammography images [24]. At first, gabor wavelet transform is used to get the feature vector pertaining to each mammography image. Then, several experiments were done for performing tenfold cross validation and analyzing the data complexity on each fold. Decision-making stage from machine learning methods is applied and attained above 0.939 of mean accuracy, mean sensitivities as high as 0.951, and greater than 0.92 as mean specificities.

Hassanien, Aboul Ella [25] introduced a scheme that combines the advantages of pulse coupled neural networks (PCNNs) and support vector machine, in conjunction with type-II fuzzy sets and wavelet to detect breast cancer on a group of MRI images. The region of interest ROI and the boundary of the breast pattern can be detected by applying a type-II fuzzy and PCNN in which help to enhance the contrast of the input image,. The experimental results of this work showed that SVM has been attained by 0.98% accuracy.

An ensemble classifier method called Population Rescaled Differential Evolution with Weighted Boosting (PRDE-WB) are used for early breast cancer detection by Jeyanthi, K [21]. At first, to enhance the early breast cancer detection, Logarithmic Cube-root Shift Preprocessing technique is initially applied to the input breast images to improve the contrast of the given input image. Next, Population Rescaling Factor with the PRDE optimization technique is used to exactly sketching the contour. Finally, PRDE is implemented with Weighted Boosting that performs early breast cancer detection using MAXMIN Ensemble Classifier algorithm. The suggested PRDE-WB method exceeds other existing methods in terms of Peak Signal-to-Noise Ratio (PSNR), overall classification accuracy, early breast cancer detection rate and early breast

cancer detection time. The Peak Signal-to-Noise rate is diminished by 22% while the time consumed for early detection of breast cancer is minimized by 15%. The overall classification achieves 18% accuracy and the rate of early breast cancer detection has 30% accuracy as compared to state-of-the-art works.

Some of machine learning algorithms were used by Osareh Alireza and Bita Shadgar [19] . The proposed mechanism was able to distinguish between benign and malignant breast cancer by using support vector machines, K-nearest neighbors and probabilistic neural networks classifiers in which each one was combined with signal-to-noise ratio feature ranking, sequential forward selection-based feature selection and principal component analysis feature extraction. Using two widely used breast cancer benchmark datasets, support vector machines give the best overall accuracy for detecting breast cancer with values equal to 98.80% and 96.33% respectively.

Sadhukhan, al in [7] used machine learning to train a model by using the predicted features of the nuclei of cells. As well as, A comparative study of two different algorithms KNN and SVM are conducted where the accuracy of each classifier is measured. KNN was achieved a better accuracy 97.48 compare to 97.27 with SVM. Besides, image processing was used to analyze a digital image of a fine needle aspirate (FNA) of breast tissue to discover the features of nuclei of the cells, and then, the feature values were applied to the trained model to find whether the tumor developed is benign or malignant.

Wan, Nathan, al in [8] worked on alternative ways to screen tests based on cell-free DNA (cfDNA). Plasma samples (N = 546 colorectal cancer and 271 non-cancer controls) were used and whole-genome sequencing was applied to extract cfDNA. IchorCNA was performed to assess cfDNA tumor fraction. General performance was estimated by using trained models through kfold cross-validation and confounder-based cross-validations methods. In a colorectal cancer cohort, a mean AUC of 0.92 (95% CI 0.91–0.93) with a mean sensitivity of 85% (95% CI 83–86%) at 85% specificity was achieved. In tumor stage and increasing tumor fraction, Sensitivity generally was raised. The validation of the machine learning methods used is underway.

The performance of the Breast Cancer Risk Assessment Tool (BCRAT) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) models was compared with the performance of eight different machine learning methods by Ming,

Chang [9]. Eight simulated datasets and two retrospective samples were used - a random population-based sample of U.S. breast cancer patients and their cancer-free female relatives (N = 1143), and a clinical sample of Swiss breast cancer patients and cancer-free women seeking genetic evaluation and/or testing (N = 2481). Adaptive Boosting of the ML method achieved 88.28% accuracy, and 88.89% using ML-random forest versus 62.40% with BCRAT for the U.S. population-based sample. 90.17% predictive accuracy was reached using ML-adaptive boosting, 89.32% using ML-Markov chain Monte Carlo generalized linear mixed model, and 59.31% with BOADICEA for the Swiss clinic-based sample.

Blanco, Jose Liñares, al in [6] used Machine Learning-based approach for gene expression analysis using RNA-seq data for cancer research within the R framework and to compare it with a classical gene expression analysis approach. Two models were proposed, a standard statistical analysis of a set of gene expression data and secondly an approach based on automated learning algorithms and techniques.

Abdel-Ilah, Layla [10], suggested a way in Artificial Neural Network (ANN) to enhance the accuracy of estimation of breast cancer cases to be malignant or benign by selecting the number of hidden layers, the number of neurons in the hidden layer and the type of activation functions in hidden layers. Wisconsin Breast Cancer Database (WBCD) was used as a sample to train the network. Three transfer functions: LOGSIG, TANSIG, and PURELINE were applied with mean square error (MSE). The performance for the different numbers of neurons in the hidden layer (20, 21, 22, 23, 24 neurons) was conducted. The best network design was achieved using 21 neurons in the hidden layer, and TANSIG as the activation function.

Abdel-Zaher, Ahmed M and Ayman M. Eldeib [11] develop a scheme to detect breast cancer using deep belief network unsupervised path followed by back-propagation supervised path. The suggested network uses a back-propagation neural network with Liebenberg Marquardt learning function, while, the weights of the network was initialized by the deep belief network path (DBN-NN). The Wisconsin Breast Cancer Dataset (WBCD) used as input to the network. The proposed work gives an accuracy of 99.68% indicating promising results over previously-published studies.

Several machine learning methods were applied to classify a bladder cancer (BC) by Adam, Liana, et al [12]. Some of the used methods are, random forests of classification trees, nearest shrunken centroids, and regularized logistic regression. The proposed worked was able to achieved to 89% predictive accuracy for detecting the presence or absence of BC by using a Logistic regression modeling, 92% accuracy to differentiate between invasive BC from other cases, 100% accuracy for recognizing muscle invasive BC (MIBC) from controls, and 79% accuracy for three-way classification between MIBC, non-muscle-invasive BC (NIMBC), and controls.

K-nearest Neighbor Machine Learning Algorithm was used to detect breast cancer by Alarabeyyat, Abdulsalam, and Mohannad Alhanahnah [13]. The proposed algorithm was divided into two parts, in the first one, features and patterns were extracted from mammography images using image process techniques. In the second part, two algorithms of the supervised learning models were used, which are the Back Propagation Neural Network (BPNN) model and the Logistic Regression (LR). Also, the input data to these algorithms are the extracted features from mammography images. The proposed model has a good regression value using BPNN that exceeded 93% with only 240 features.

Bazazeh, Dana, and Raed Shubair use machine learning (ML) as an effective mechanism for early detection and diagnosis of breast cancer. This work compares three popular methods, namely Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN). The Wisconsin original breast cancer data set was used as a training set to evaluate and compare the performance of the three ML classifiers in terms of key parameters such as accuracy, recall, precision. The highest performance was gain by using the SVMs algorithm.

Bhat, Junaid Ahmad, Vinai George, and Bilal Malik [15] used adaptive resonance theory to help in early diagnosis of breast cancer. This project uses a cloud computer to help a clinician to use the system for the early diagnosis of the patients using a web-based interface from anywhere in the world. Wisconsin dataset was used. Several tests have been taken and the better results can be achieved by taking the vigilance parameter as 0.5 and taking the ratio of data as 90% for training and 10 % for testing.

Hu, Zilong, et al. [17] provides a survey on the applications of deep learning for cancer detection and diagnosis. Convolutional neural networks, fully convolutional networks, autoencoders, and deep belief networks were reviewed. Different cancer types were studied and diagnosis by the reviewed applications: lung cancer, skin cancer, prostate cancer, brain cancer, colonial cancer, cervical cancer, bladder cancer, and liver cancer.

Jongerius, Chiara [51] conducted a general review to show mobile health (mHealth) apps have been utilized to advance prevention, improve early detection, monitor care, and provide survivors and patients with necessary information they need. In this paper, the available research-tested interventions using mHealth apps were reviewed to show their impact on breast cancer care. The writer performed a systematic search of Medline, PsycINFO, Embase, and Scopus to include the relevant studies and found 29 empirical ones that provide a health care for breast cancer. However, the utility of mHealth apps are still limited and just used for feedbacks after some time but not in real time and alarms patients what should to do next time even if before surgery or after.

Transpara [52,53] is an AI computer system used by radiologists to detect breast cancer and is a good tool for decision making to help in the diagnosis of cancer. The system works on mammography images and compatible with examinations done in digital mammography and/or breast tomosynthesis systems from different vendors. Deep learning convolutional neural networks technique and features classifiers and image analysis algorithms are used by this system to detect lesions of the breast. The system is able to determine suspicious region findings from soft tissue and calcification at images. More than one million trusted mammography images are collected from different databases to train the system, therefore; the overall accuracy of this system is superior to others that already exist according to the writer. This system is not free of charge, as well as suggesting the suitable treatment is not taken into consideration. Besides, the suggested application is only used by radiologists as supporting tool to help at early detection of cancer. Using the application for learning purposes does not take into consideration and this important point to support students or teachers at faculty of medicine with a tool to provide them with ability to detect and diagnose breast cancer and helps to choose the suitable treatments. This application (desktop application) is built using server-client architecture and is not a website, therefore, the maintenance process will be painful.

Significance of the study

The significance of the study is that it intends to explore the most accurate machine learning algorithms for detecting and/or diagnosing breast cancer in order to create a complete system which will be able to suggest the suitable treatment to increase the survival rate. Besides, it will improve the skills required for doctors, radiologists, students of colleges, and IT or computer science researchers by using the proposed system that will provide them with different breast cancer cases.

Problem Statement

Breast cancer is one of the most frequent cancer in women that causes death. Generally, death cases of breast cancer are increasing notably all over the world. Therefore, early detection of breast cancer is very important issue that can result in lessening treatment by radiation and increase the survival rate. A few of modalities are used to help radiologists to detect and diagnose breast cancer more accurately and efficiently such as Ultrasound, mammography, and magnetic resonance modalities. The researchers suggested different ways not only to detect the cancer and draw its location on medical images but also to determine the type of a lesion is benign or cancer. However, there is no comprehensive research that detects and diagnoses and suggests the suitable treatments. Simultaneously, many studies about breast cancer detection are conducted using data mining, image processing, or machine learning, however, the most accurate results were achieved by applying the techniques of machine learning. To improve the detection and diagnosis of breast cancer, a large number of annotated medical images with different cases should be used to train the models of machine learning, but most of the datasets exist without any annotations. Therefore, creating a complete system to deal with datasets, even if it has annotated images or not, is a challenging task. Besides, suggesting a suitable treatment for each image inside the collected dataset is a hard problem since the right treatment or drugs should be revised by the experts of the domain.

Objectives

- To study various existing techniques in machine learning algorithms and face the challenges posed by breast cancer detection.
- To collect as much as cancer datasets to be enough and able to detect breast cancer and suggest a suitable treatment.
- To study and identify the required features of datasets to be able to implement in this system.
- To determine appropriate methods that can enhance cancer detection challenges.
- To apply the techniques of multimedia on the suggested system to create interactive environment easy to be accessed and used.
- To test the output of the final product with new medical images even annotated or not.
- Get final feedback by radiologists or doctors about how accurate the outcomes of the application are.

Motivation

At first, what inspired me in this area is my work done in master project, which was mainly to collect different types of medical images of breast cancer at a global archive and allows a radiologist to anonymize and annotate the images. Therefore, in order to complete the suggested master project, a complement project should be added, which leads to a final system with ability to not only detect and diagnose breast cancer but also suggest the right treatment. Secondly, There are many registered death cases because misusing of drugs or wrong diagnosis of breast cancer by doctors or radiologists with limited knowledge of breast cancer, leads to death in most cases, especially in my country Yemen. Therefore, to increase the number of survival, there is a basic need of a complete system to deal with different types of images that are issued from different modalities, to detect breast cancer cases, and to suggest a suitable treatment. As well as, radiologists or doctors can increase their knowledge by reviewing different breast cancer cases with suitable treatment or drugs too. Besides, collecting different datasets with suitable treatment and revising the suggest treatment will enhance the accuracy of detecting and diagnosing of breast cancer. Finally, students at faculty of medicine need many medical

images of breast cancer with suitable treatment to enhance their skills in detection and diagnosis.

Gap

In this part, the reviewed papers can be divided into two groups: first one contains papers which focusing on applying just algorithms to detect or diagnose breast cancer; the second one also have papers with algorithms but the only difference is these algorithms are used in real applications (mobile application, websites, or desktop applications).

At first group, many work have been published at very popular magazine such as Springer, Science Direct (SD), IEEE Xplore Digital Library, and PubMed, Scopus, and EBSCO, ACM, and Elsevier. They used different algorithms to detect and diagnose breast cancer; achieved high values in accuracy. Most of the proposed algorithms only were focusing in just detection and diagnosis of breast cancer while a suitable treatment for each case of breast cancer didn't taken into their accounts. To detect and diagnose breast cancer, physicians or radiologists have been used different images that are issued by using a special device called modality responsible to screen any organ of human body like brain and breast. These modalities are Digital Mammograms (DM), Ultrasound (US), Magnetic Resonance Imaging (MRI), Microscopic (Histological) images, Infrared Thermography, Microwave; and each device issues only one type of images. Almost a big number of the reviewed papers just deal with one type of image. Few papers deal with two type of images but not all type images that are used in breast cancer. Different techniques already exist to detect breast cancer like soft computing and simulation techniques but the most accurate results were gained by machine learning techniques. According to the proposed works at the collected papers; many methods are used like logistic regression, k-Nearest-Neighbor (KNN), Support vector machine (SVM), naïve Bayes, decision trees, convolutional neural network, sparse autoencoder, stacked sparse autoencoder, Graph-based Semi-supervised Learning (SSL), Bayesian Networks (BN), Decision Tree (DT), and SSL Co-training methods, but the SVM is superior the others methods and achieves the highest accurate outcomes and recommended by many papers. Another writers like Sanwal, Karan, and Himanshu Ahuja [37] focus on how to reduce the computational expense in the diagnosis of breast cancer, therefore; a novel ensembling method is proposed, namely CFDNN(Breast Cancer

Detection Using Low-Computation-Based Collaborating Forward-Dependent Neural Networks. Most of the writer used popular dataset such as Wisconsin Breast Cancer Dataset (WBCD), Mini Mammographic Image Analysis Society(Mini-MIAS) dataset, digital mammography (full-field digital mammography (FFDM), and SEER cancer database, to test their proposed methods. Besides, Windows and Linux platforms used by writers of these papers and their proposed methods were implemented using Matlab, R framework, Python, and Java.

At second group, many papers have been published to early detect and diagnose breast cancer. Some writers had implemented their proposed project based on web interface and others on mobile applications. Bhat, Junaid Ahmad, Vinai George, and Bilal Malik [15] used adaptive resonance theory to help in early diagnosis of breast cancer. This project uses a cloud computer to help a clinician to use the system for the early diagnosis of the patients using a web-based interface from anywhere in the world. The Wisconsin dataset was used. But this work only focused on detection of breast cancer. Jongerius, Chiara [51] conducted a general review to show mobile health (mHealth) apps have been utilized to advance prevention, improve early detection, monitor care, and provide survivors and patients with necessary information they need. In this paper, the available research-tested interventions using mHealth apps were reviewed to show their impact on breast cancer care. The writer performed a systematic search of Medline, PsycINFO, Embase, and Scopus to include the relevant studies and found 29 empirical ones that provide a health care for breast cancer. However, the utility of mHealth apps are still limited and just used for feedbacks after some time but not in real time and alarms patients what should to do next time even if before surgery or after.

Drawbacks

Based on the reviewed papers mentioned above on the gap section, many drawbacks are noticed. In the first group of papers, most algorithms are not applied in real medical application to help doctors or patients to detect and diagnose breast cancer. As well as, all these papers did not provide any way to suggest a suitable treatment for each image of breast cancer. Therefore, there is still a place for improvement. First, a dataset with millions of annotated trusted images is required to enhance the output of deep learning methods. Second, a complete system using the

techniques of multimedia to be used for learning purposes still not exist or exist with a lot of limitations, therefore, the learners do not get a complete benefit to learn different breast cancer cases with real images and diagnose breast cancer and suggest the treatment. Third, it is necessary to develop techniques to efficiently detect and diagnose cancer using images acquired from different scanning modalities not only one or two like the reviewed papers above.

Datasets

A few of standard datasets are used in medical areas some have annotated by radialogisrs and others have not like: CBIS-DDSM, Wisconsin Breast Cancer Dataset (WBCD), Mini Mammographic Image Analysis Society(Mini-MIAS), INbreast, and SEER cancer databases. CBIS-DDSM (Curated Breast Imaging Subset of DDSM) is an updated and standardized version of the Digital Database for Screening Mammography (DDSM) [54], a database includes 2,620 scanned film mammography. In addition to, each image contains normal, benign, and malignant cases with verified pathology information. Wisconsin Breast Cancer Dataset (WBCD) contains 699 images and was obtained from the University of Wisconsin Hospitals [55]. The Digital Database for Screening Mammography (DDSM) contains approximately 2,500 mammographic images while MIAS Mini-Mammographic Database (i.e. mini-MIAS database of mammograms) has 322 digitized films [56]. Digital mammography also called full-field digital mammography (FFDM), is a mammography system in which the X-ray film is replaced by solid-state detectors that convert X-rays into electrical signals. The INbreast database contains a total of 115 cases (410 images) of which 90 cases are from women with both breasts (4 images per case) and 25 cases are from mastectomy patients (2 images per case) [58]. SEER collects a complete and accurate data not only for breast cancer but also others cancers, besides, it contains thousands of images [60].

Advantages of the proposed application

According to the above sections on gap and drawbacks sections, it is beneficial for doctors and also patients to create a complete system to not only detect and diagnose breast

cancer but also advice the suitable treatment. To enhance the proposed application, many advantages of existing system and papers will be collected in one application and the drawbacks will be waived. Therefore, the first step is to survey deeply the best existing system and the best accurate algorithms to be implemented in this application or if it is necessary to enhance the accuracy of the best algorithms already done before. Second, after surveying we will analysis the reviewed work to determine which will fit our needs. The proposed system will focus on accuracy and reduce the computational expense in the detection and diagnosis of breast cancer. Besides, this system can be used for learning purposed by students at faculty of medicine, the techniques of multimedia will be used to make it user-friendly and different from other system already exist.

Multimedia

Multimedia is a class of computer-driven interactive communication systems which create ,store, transmit, and retrieve textual, graphic, and auditory networks of information [63]. It may refers to the use of several different media to convey information (text, audio, graphics, animation, video, and interactivity). Multimedia is applied mostly in games, which contains from texts, audios, images and animations. The techniques of multimedia made games more innovative and interactive. Besides, it is also used to give special effects in movies and animations. Therefore, implemented such techniques will for sure make the final system more innovative and interactive for users especially if the system will be created for learning purposes.

Tools and software

This application can be implemented as a website on Windows or Linux platform according to the selected programming language like MVC, Asp.net, PHP, or Python. However, it possible to use more one programming language like using Matlab to analysis the outputs of some of algorithms and draws some of figures for each case; Asp.net can be used to collected the datasets and the images of breast cancer. Therefore, the most suitable tools and programming language will be determined after the process of analyzing the current algorithms and systems.

Platforms

It will not be a problem to determine the best operating system that can be used even if it is Windows or Linux platform since the suggested application is a website and it depends on the chosen programming language, in short, it can work on any platform. However, the adequate platform will be determined later.

Methodology

The objectives of the research can be accomplished and explained in this section. The system will be divided into three phases: training and testing phases, and the last phase is to apply the techniques of multimedia on the proposed system. The training phase is the stage in which the system will collect a group of well-known datasets to be used as input to train the system model to produce a predictive model as shown in figure 1. Figure 2 shows the second phase of the proposed system where unknown images will be checked by the system to detect breast cancer and suggest a suitable treatment.

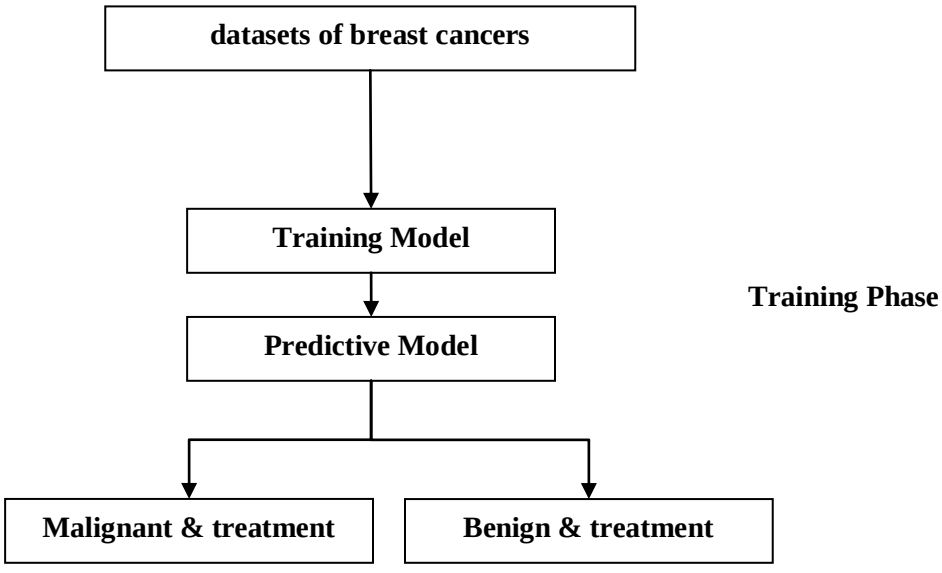

Figure 1: the suggested methodology for training phase.

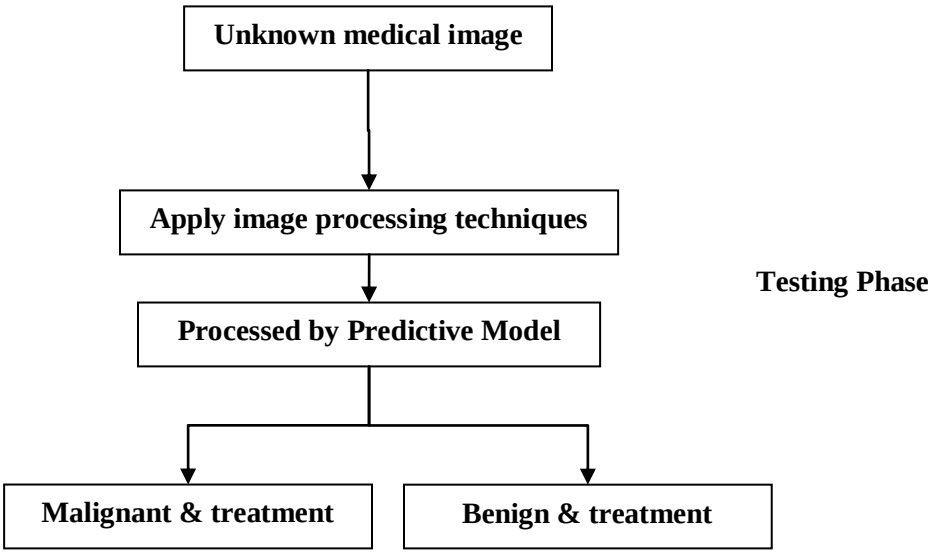

Figure 2: the suggested methodology for testing phase.

- **Training Phase**

1. **Datasets of breast cancer:** There are several standard datasets used throughout recent years that can be used in this system some of them have ROIs annotations and others are not.
2. **Training model:** In this phase, we will choose the best algorithm already used to detect breast cancer to train the system model using the collected datasets.
3. **Predictive Model:** In this phase, the system will be able to detect the cancer type (Malignant or Benign) and also suggest a suitable treatment.

- **Testing Phase**

1. **Unknown medical image:** an image that can be a Digital Imaging and Communication in Medicine (DICOM) file or a normal image file (like jpg, tiff, …ect).
2. **Applying image processing techniques:** different techniques can be used to get the necessary features from images.

3. **Processed by Predictive Model:** in this case, the unknown image will be processed by the predictive model to get the right prediction of breast cancer and show suitable treatment to a user**.**

- **Applying the multimedia techniques**

  The techniques of multimedia should be taken into consideration in this system to create a friendly system for users.

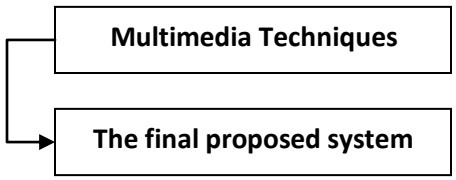

Figure 3: applying the techniques of multimedia.

Extra phase might be studied, validates the outcomes of the system by getting the final feedback from radiologists or doctors and correct the mistaken outputs or any errors that appear in the whole life cycle of the proposed system in which will lead to negative results.

Conclusion

Recent enhancements in the evaluation of breast cancer detection algorithms are mentioned in this proposal according to research conducted by various researchers. Even though all the approaches suggested by different researchers differed in their own way, there is a commonality of adopting knowledge from the same dataset used in previous studies. Since each proposal work with one or more algorithms, no complete research detect breast cancer and suggest a suitable treatment at the same time. The diversity of the proposed algorithms to detect breast cancer makes a challenge to choose the best one or to enhance one of them to get a better outcome. Some of the algorithms can work well for some cases of breast cancer but not for all breast cancer images. There are two datasets groups already exist: the first one contains annotated images and the second contains images without any annotations, therefore, find a way to work with both groups to achieve good performance is a hard task. Collecting many datasets

for breast cancer is not an easy task since the collected datasets should have some features in common to be able to be detected through one application, this is from one hand. On another hand, if some required features are missing, it takes a time to add these features to the collected datasets from experts of the domain. Finding a unique application to do all the suggested work and be able to deal with different types of images (DICOM, jpg, tiff, etc) will be a hard task since each type of image needs a different way to read or extract some required features from it. The system will be used for first stage for learning purposes even if at the faculty of medicine by teachers or students, or at a laboratory by radiologists who have small experience in detection and diagnosis of breast cancer.

References

[1] Deo, Rahul C. "Machine learning in medicine." Circulation 132.20 (2015): 1920-1930.

[2] Shen, Li, et al. "Deep Learning to Improve Breast Cancer Detection on Screening Mammography." Scientific reports 9.1 (2019): 1-12.

[3] Al-Tam, Riyadh Mohammed. Diversifying medical imaging of breast lesions. Diss. 2015.

[4] https://towardsdatascience.com/understanding-cancer-using-machine-learning-84087258ee18

[5] Abu-Mostafa, Yaser S., Malik Magdon-Ismail, and Hsuan-Tien Lin. Learning from data. Vol. 4. New York, NY, USA:: AMLBook, 2012.

[6] Blanco, Jose Liñares, et al. "Differential Gene Expression Analysis of RNA-seq Data Using Machine Learning for Cancer Research." Machine Learning Paradigms. Springer, Cham, 2019. 27-65.

[7] Sadhukhan, Subham, Nityasree Upadhyay, and Prerana Chakraborty. "Breast Cancer Diagnosis Using Image Processing and Machine Learning." Emerging Technology in Modelling and Graphics. Springer, Singapore, 2020. 113-127.

[8] Wan, Nathan, et al. "Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA." BMC cancer 19.1 (2019): 832.

[9] Ming, Chang, et al. "Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models." Breast Cancer Research 21.1 (2019): 75.

[10] Abdel-Ilah, Layla, and Hana Šahinbegović. "Using machine learning tool in classification of breast cancer." CMBEBIH 2017. Springer, Singapore, 2017. 3-8.

[11] Abdel-Zaher, Ahmed M., and Ayman M. Eldeib. "Breast cancer classification using deep belief networks." Expert Systems with Applications 46 (2016): 139-144.

[12] Adam, Liana, et al. "Plasma microRNA profiles for bladder cancer detection." Urologic Oncology: Seminars and Original Investigations. Vol. 31. No. 8. Elsevier, 2013.

[13] Alarabeyyat, Abdulsalam, and Mohannad Alhanahnah. "Breast cancer detection using k-nearest neighbor machine learning algorithm." 2016 9th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2016.

[14] Bazazeh, Dana, and Raed Shubair. "Comparative study of machine learning algorithms for breast cancer detection and diagnosis." 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA). IEEE, 2016.

[15] Bhat, Junaid Ahmad, Vinai George, and Bilal Malik. "Cloud computing with machine learning could help us in the early diagnosis of breast cancer." 2015 Second International Conference on Advances in Computing and Communication Engineering. IEEE, 2015.

[16] Hornbrook, Mark C., et al. "Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data." Digestive diseases and sciences 62.10 (2017): 2719-2727.

[17] Hu, Zilong, et al. "Deep learning for image-based cancer detection and diagnosis− a survey." Pattern Recognition 83 (2018): 134-149.

[18] Kumar, Aman, and M. Poonkodi. "Comparative Study of Different Machine Learning Models for Breast Cancer Diagnosis." Innovations in Soft Computing and Information Technology. Springer, Singapore, 2019. 17-25.

[19] Osareh, Alireza, and Bita Shadgar. "Machine learning techniques to diagnose breast cancer." 2010 5th International Symposium on Health Informatics and Bioinformatics. IEEE, 2010.

[20] Guzmán-Cabrera, R., et al. "Digital image processing technique for breast cancer detection." International Journal of Thermophysics 34.8-9 (2013): 1519-1531.

[21] Jeyanthi, K., and S. Mangai. "Ensembled Population Rescaled Differential Evolution with Weighted Boosting for Early Breast Cancer Detection." Mobile Networks and Applications (2019): 1-15.

[22] Aydin, Emine Avşar, and Mümine Kaya Keleş. "UWB Rectangular Microstrip Patch Antenna Design in Matching Liquid and Evaluating the Classification Accuracy in Data Mining Using Random Forest Algorithm for Breast Cancer Detection with Microwave." Journal of Electrical Engineering & Technology (2019): 1-10.

[23] Aydin, Emine Avşar, and Mümine Kaya Keleş. "UWB Rectangular Microstrip Patch Antenna Design in Matching Liquid and Evaluating the Classification Accuracy in Data Mining Using Random Forest Algorithm for Breast Cancer Detection with Microwave." Journal of Electrical Engineering & Technology (2019): 1-10.

[24] Ghasemzadeh, Ardalan, Saeed Sarbazi Azad, and Elham Esmaeili. "Breast cancer detection based on Gabor-wavelet transform and machine learning methods." International Journal of Machine Learning and Cybernetics 10.7 (2019): 1603-1612.

[25] Hassanien, Aboul Ella, et al. "Breast cancer detection and classification using support vector machines and pulse coupled neural network." Proceedings of the Third International Conference on Intelligent Human Computer Interaction (IHCI 2011), Prague, Czech Republic, August, 2011. Springer, Berlin, Heidelberg, 2013.

[26] Jalalian, Afsaneh, et al. "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review." Clinical imaging 37.3 (2013): 420-426.

[27] R.Chtihrakkannan, P.Kavitha, T.Mangayarkarasi, R.Karthikeyan. " Breast Cancer Detection using Machine Learning." International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September 2019.

[28] Keatmanee, Chadaporn, et al. "Automatic initialization for active contour model in breast cancer detection utilizing conventional ultrasound and Color Doppler." 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017.

[29] Khan, A. A., and A. S. Arora. "Breast Cancer Detection Through Gabor Filter Based Texture Features Using Thermograms Images." 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). IEEE, 2018.

[30] Khourdifi, Youness, and Mohamed Bahaj. "Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification." 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS). IEEE, 2018.

[31] Kontos, Konstantinos, and Manolis Maragoudakis. "Breast cancer detection in mammogram medical images with data mining techniques." IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, Berlin, Heidelberg, 2013.

[32] Kourou, Konstantina, et al. "Machine learning applications in cancer prognosis and prediction." Computational and structural biotechnology journal 13 (2015): 8-17.

[33] Král, Pavel, and Ladislav Lenc. "LBP features for breast cancer detection." 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016.

[34] Lu, Yao, et al. "A Review of Breast Cancer Detection in Medical Images." 2018 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2018.

[35] Lu, Hao-Chun, El-Wui Loh, and Shih-Chen Huang. "The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection." 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2019.

[36] Michahial, Stafford, and Bindu A. Thomas. "Applying cuckoo search based algorithm and hybrid based neural classifier for breast cancer detection using ultrasound images." Evolutionary Intelligence (2019): 1-18.

[37] Sanwal, Karan, and Himanshu Ahuja. "Breast Cancer Detection Using Low-Computation-Based Collaborating Forward-Dependent Neural Networks." Advances in Data and Information Sciences. Springer, Singapore, 2019. 105-115.

[38] Selvathi, D., and A. Aarthy Poornila. "Deep learning techniques for breast cancer detection using medical image analysis." Biologically Rationalized Computing Techniques For Image Processing Applications. Springer, Cham, 2018. 159-186.

[39] Sharma, Shubham, Archit Aggarwal, and Tanupriya Choudhury. "Breast Cancer Detection Using Machine Learning Algorithms." 2018 International Conference on

Computational Techniques, Electronics and Mechanical Systems (CTEMS). IEEE, 2018.

[40] Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep learning in medical image analysis." Annual review of biomedical engineering 19 (2017): 221-248.

[41] Song, Hongchao, Yunpeng Li, and Aidong Men. "Microwave breast cancer detection using time-frequency representations." Medical & biological engineering & computing 56.4 (2018): 571-582.

[42] Suresh, Veena, and R. Kiran. "3D Inverse Electromagnetic Solver Using Deep Neural Network Towards Breast Cancer Detection." 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, 2018.

[43] Syed, Liyakathunisa, Saima Jabeen, and S. Manimala. "Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques." Advances in Soft Computing and Machine Learning in Image Processing. Springer, Cham, 2018. 149-183.

[44] Takada, Masahiro, et al. "Prediction of postoperative disease-free survival and brain metastasis for HER2-positive breast cancer patients treated with neoadjuvant chemotherapy plus trastuzumab using a machine learning algorithm." Breast cancer research and treatment 172.3 (2018): 611-618.

[45] Wang, Xin, et al. "A Nomogram To Predict The Overall Survival Of Breast Cancer Patients And Guide The Postoperative Adjuvant Chemotherapy In China." Cancer Management and Research 11 (2019): 10029.

[46] Uzer, Mustafa Serter, Onur Inan, and Nihat Yılmaz. "A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA." Neural Computing and Applications 23.3-4 (2013): 719-728.

[47] Vasquez, JA Tobon, et al. "Microwave imaging for early breast cancer detection: Experimental testing of a low-cost portable system." 2016 IEEE International Symposium on Antennas and Propagation (APSURSI). IEEE, 2016.

[48] Yassin, Nisreen IR, et al. "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review." Computer methods and programs in biomedicine 156 (2018): 25-45.

[49] Shirazi, Amin Zadeh, Seyyed Javad Seyyed Mahdavi Chabok, and Zahra Mohammadi. "A novel and reliable computational intelligence system for breast cancer detection." Medical & biological engineering & computing 56.5 (2018): 721-732.

[50] Modiri, Arezoo, et al. "Review of breast screening: Toward clinical realization of microwave imaging." Medical physics 44.12 (2017): e446-e458.

[51] Jongerius, Chiara, et al. "Tested Mobile Apps for Breast Cancer Care: Systematic Review." JMIR mHealth and uHealth 7.2 (2019): e10930.

[52] Rodríguez-Ruiz, Alejandro, et al. "Detection of breast cancer with mammography: effect of an artificial intelligence support system." Radiology 290.2 (2019): 305-314.

[53] Rodriguez-Ruiz, Alejandro, et al. "Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists." JNCI: Journal of the National Cancer Institute 111.9 (2019): 916-922.

[54] Nolan, T. (2019). CBIS-DDSM - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. [online] Wiki.cancerimagingarchive.net. Available at: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM [Accessed 21 Feb. 2020].

[55] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. Available at: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)

[56] Grgic, M. (2020). Mammographic Image Analysis Homepage - Databases. [online] Mammoimage.org. Available at: http://www.mammoimage.org/databases/ [Accessed 21 Feb. 2020].

[57] Eng.usf.edu. (2020). USF Digital Mammography Home Page. [online] Available at: http://www.eng.usf.edu/cvprg/Mammography/Database.html [Accessed 21 Feb. 2020].

[58] Frias, R. (2020). Get INbreast Database - Breast Research Group. [online] Medicalresearch.inescporto.pt. Available at: http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database [Accessed 21 Feb. 2020].

[59] Frias, R. (2020). Projects - Breast Research Group. [online] Medicalresearch.inescporto.pt. Available at: http://medicalresearch.inescporto.pt/breastresearch/index.php/Projects [Accessed 21 Feb. 2020].

[60] SEER. (2020). Accessing the Data - SEER Datasets. [online] Available at: https://seer.cancer.gov/data/access.html [Accessed 21 Feb. 2020].

[61] Gayeski, Diane Mary. Multimedia for learning: Development, application, evaluation. Educational Technology, 1993.

Research Student **Research Guide**

Mr. Riyadh Mohammed Musaed Al-Tam **Dr. Sachin M. Narangale**